

A LEVEL-SET APPROACH FOR TRACKING OBJECTS IN IMAGE SEQUENCES USING A LEVEL CONSERVATION CONSTRAINT: APPLICATION TO CARDIAC SEQUENCES

T. Dietenbeck^{1,2}, D. Barbosa^{2,3}, M. Alessandrini³, J. D'hooge³, D. Friboulet², O. Bernard²

¹ Clermont Université, Université d'Auvergne, ISIT, BP10448, F-63000 Clermont-Ferrand.
CNRS, UMR6284, BP10448, F-63000 Clermont-Ferrand.

² Université de Lyon, CREATIS; CNRS UMR5220; Inserm U1044;
INSA-Lyon; Université Lyon 1, France

³ Cardiovascular Imaging and Dynamics, Katholieke Universiteit Leuven, Leuven, Belgium

ABSTRACT

Tracking of moving objects in an image sequence is an important task in many application (*e.g.* medical imaging, robotics). However, this task is usually difficult due to inherent problems that could happen in sequences (*i.e.* possible occlusion of the object, large interframe motion). In this paper, we describe a new approach to integrate *a priori* motion information into a level-set-based active contour approach. Specifically, we introduce a new constraint that enforces the conservation of the levels of the implicit function along the image sequence. This constraint is formulated as a motion prior energy and is used in a tracking algorithm. The method is validated quantitatively on a clinical application based on 10 echocardiographic and 5 cine-MRI sequences (≈ 700 images).

Index Terms— Segmentation, tracking, active contour, level-set, shape prior, motion prior, cardiac imaging

1. INTRODUCTION

The problem of tracking moving objects in video sequences remains of great interest in the framework of various applications (*e.g.* medical imaging, video surveillance, robotics). When segmenting sequences and tracking an object over time, the knowledge of the underlying motion may bring valuable information and help improving segmentation results and speed. Indeed, parts of the object that are occluded in a frame might be visible in another one while temporal coherence implies that contours in successive frames have similar shapes.

Variational approaches and region-based active contours have proven to be efficient for tracking purposes. In this context, some authors proposed to integrate motion either directly into the snake formalism [1] or through Kalman filtering [2]. However snakes methods have the drawback of requiring a regridding mechanisms and do not handle topological changes or region terms in the minimized energy. We thus choose to focus on level-set techniques, where tracking methods can be classified broadly into two categories: motion-based approaches [3, 4] or model-based approaches [5, 6].

The motion-based approaches either assume the motion to be estimated prior to the tracking [4] or perform a joint motion estimation and segmentation [3]. The methods have complementary advantages and drawbacks: while relying on *a priori* estimated motion allows selecting the best motion estimation algorithm independently from the segmentation approach, the joint approach allows the segmentation step to improve motion estimation and vice-versa. On the other hand, model-based methods use prior knowledge which is included either through a dynamical shape prior [6] or by constraining the contour evolution between consecutive frames [5]. This allows modeling any kind of motion but may require a learning step.

In this article, we propose a new motion prior energy criterion whose minimization enforces the conservation of the zero-level of the implicit function along the images sequence. We then derive this energy to obtain an evolution term which acts as a motion prior. The interest of such an approach resides in the fact that the level-set is forced to move in accordance to the estimated motion while the aforementioned methods try to separate the image in region of homogeneous motion. The paper is organized as follows. In Section 2, we recall the general level-set framework. In Section 3, we describe the level conservation constraint and we show how to embed it into a motion prior energy. Implementation issues are discussed in Section 4. Experimental results on echocardiographic and cine-MRI sequences with clinical interest are presented in Section 5. We give the main conclusions and perspectives in Section 6.

2. CONTEXT

2.1. Level-set framework

Let $\Omega \in \mathbb{R}^2$ denotes the image space. In the level-set formalism, the evolving interface $\Gamma \in \mathbb{R}^2$ is represented as the zero level-set of a Lipschitz-continuous function $\phi : \Omega \rightarrow \mathbb{R}$. The problem of segmenting one object from the background is then handled by the evolution of one level-set driven by the minimization of a specific energy criterion; its steady state

partitions the image into two regions that delimit the boundaries of the object to be segmented [7].

2.2. Motion estimation

We chose to use the motion estimation technique recently proposed by Alessandrini *et al.* [8]. It is based on the monogenic signal which extends the concept of analytic signal to multiple dimensions. The brightness consistency assumption usually used in motion estimation (Optical Flow) is replaced by a monogenic phase consistency which has proven to be more robust. The authors proposed a multi-scale approach in order to be able to estimate large motion. The motion is assumed to be locally affine and estimated for several neighborhood sizes using the monogenic phase. The retained displacement estimate then corresponds to the one with the smallest residual error. More details on this approach can be found in [8].

3. OUR CONTRIBUTION: MOTION PRIOR TERM

In this Section, we describe how we use the motion information to guide the evolution of the active contour. As mentioned in Section 1, this can be done in two different ways: either considering motion as a data or as a prior knowledge. In the latter case, the model is often obtained as a dynamic shape prior through a learning step (*e.g.* PCA or deep learning) which can be time consuming and requires experts to manually outline the target object in a dataset. Furthermore, one has to be very careful in the dataset construction in order not to give too much weight to a particular shape while ensuring that the training set is representative enough to cope with the variability of the shapes to be segmented. In this paper, we thus choose to consider motion as a data and assume it has been estimated prior to segmentation. Indeed this solution is flexible since at any time, we can select the current motion estimation method that produces the best results, while keeping the same segmentation algorithm.

3.1. Motion prior

The proposed motion term enforces the conservation of the level of the implicit function along the sequence. This can be written as:

$$\phi_x u + \phi_y v + \phi_t = 0 \Leftrightarrow \nabla \phi \cdot \mathbf{V} = 0, \quad (1)$$

where ϕ is the level-set, $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial t)$, $\phi_\alpha = \partial\phi/\partial\alpha$ and $\mathbf{V} = (u, v, 1)$ is the motion expressed in homogeneous coordinates.

This constraint is then expressed as an energy in a variational formalism as follows:

$$E_m = \frac{1}{2} \int_{\Omega} \delta(\phi(\mathbf{p})) (\nabla \phi(\mathbf{p}) \cdot \mathbf{V}(\mathbf{p}))^2 d\mathbf{p}, \quad (2)$$

where $\delta(\cdot)$ is the Dirac distribution allowing to select the zero level of the implicit function.

The evolution equation is obtained by minimizing (2) with respect to ϕ and is given by¹:

$$\begin{aligned} \frac{\partial \phi}{\partial \tau} = -\frac{\partial E_m}{\partial \phi} = & \delta(\phi(\mathbf{p})) (\mathbf{V}^T \mathcal{H}(\phi) \mathbf{V} \\ & + \nabla \phi^T \mathcal{J}(\mathbf{V}) \mathbf{V} + \text{Tr}(\mathcal{J}(\mathbf{V})) \nabla \phi^T \mathbf{V}) \end{aligned} \quad (3)$$

where τ is an artificial time parameter, $\mathcal{H}(\phi)$ is the Hessian matrix of ϕ , $\mathcal{J}(\mathbf{V})$ the Jacobian matrix of \mathbf{V} and $\text{Tr}(A)$ is the trace of the matrix A .

3.2. Discretization of the evolution equation

In order to avoid numerical instabilities, the derivatives of ϕ should be computed using upwind schemes [7]. Considering (3), it can be seen that we have to compute the first and second order derivatives of ϕ with respect to each dimension (x , y and t). Spatial derivatives can be easily computed though one may need to impose boundary conditions in order to compute the backward/forward derivatives on $\partial\Omega$. On the other hand, temporal derivatives (forward and 2nd order) require to use $\phi(t+1)$ which is not available when segmenting the frame t .

A first solution would be to segment the whole sequence considering a 3D level-set: $\phi : \mathbb{R}^2 \times \mathbb{R}^+ \mapsto \mathbb{R}$. However, this solution has several drawback. First, due to the local nature of the data attachment term, the initial contour on each frame has to be close to the object boundaries. As will be shown in Section 5, tracking techniques which could be used to give the initial curves, fail to segment an entire sequence properly. A second drawback is that we would need to acquire the complete sequence before being able to process it making the algorithm not usable for on-line analysis.

This issue is solved by computing a rough approximation of ϕ at t and $t+1$ from the knowledge of $\phi(t-1)$ and the available motion field: starting from the segmentation result at $t-1$, we track the object boundary using the estimated motion at $t-1$ and t to obtain a prediction of ϕ at $t+1$ which is then used for the derivatives computation.

Note that the first derivatives of the velocity \mathbf{V} are computed using standard centered derivatives. This only implies that we need to compute the motion at time $t+1$ as well.

4. IMPLEMENTATION ISSUES

We implemented our level-set evolution equation using standard finite difference scheme [7], where the implicit function is represented by a signed distance function ϕ . In order to improve efficiency, we only compute values of ϕ in a narrow band around the zero level-set. Consequently, we re-initialize ϕ every iteration using a fast marching scheme [7].

5. EXPERIMENTS

We recently proposed a method to segment the whole myocardium in echocardiographic images [9] using a local data

¹For brevity sake, the details of the derivation are not given here but can be found at http://www.creatis.insa-lyon.fr/technical-reports/LevelSet_OpticalFlow.pdf.

attachment term and a shape prior. We propose here to extend this method to the segmentation and tracking of the heart. To this end, the motion term (3) is added to the evolution equation given in [9]. The complete evolution equation of the implicit function is given by:

$$\frac{\partial \phi}{\partial \tau}(\mathbf{p}) = \nu_d f_d(\mathbf{p}) + \nu_s f_s(\mathbf{p}) + \nu_m f_m(\mathbf{p}), \quad (4)$$

where $f_{(\cdot)} = -\partial E_{(\cdot)}/\partial \phi$. f_m is given in (3), f_d is the data attachment term and f_s is the shape prior term proposed in [9].

The behavior of the algorithm is then evaluated on a dataset of 5 parasternal short-axis and 5 apical 4-chamber echocardiographic sequences acquired using a GE Vivid E9 system equipped with a 2.5 MHz M5S probe (GE Vingmed Ultrasound, Horten, Norway). Two experts then manually outlined the myocardium on one cardiac cycle per sequence yielding a dataset of approximately 600 images. We also assessed the performances of our method on 5 cine-MRI sequences of the DETERMINE study available in the Cardiac Atlas Project platform [10], where only one manual reference was available.

To evaluate segmentation results we measure the correspondence between the detected contours and the manual reference given by the experts. For US data where two reference contours were available, the mean contour was used. We adopt two different metrics, *i.e.* the Hausdorff Distance HD and the Mean Absolute Distance MAD . While the MAD measures a global correspondence between the two contours, the HD is well suited for evaluating the local behavior of the algorithm.

Our algorithm is also compared to two methods proposed recently [1, 4] that are also motion-driven active contours. Hamou *et al.* [1] proposed an algorithm close to ours since it uses both a motion and a shape term in a GVF-based snake framework and was evaluated on apical 4-chamber sequences. More specifically two edge maps are computed and used alternatively to evolve the curve: the first one is based on the optical flow computed between two subsequent frames while the second one is a shape prior based on 3^{rd} order hyperbolas. The second algorithm is a level-set motion based method described by Herbulot *et al.* [4]. It assumes the motion estimated prior to the segmentation and tries to separate the image into region with similar motion histogram distribution. Moreover, in order to perform a fair comparison, we adapted this method to the cardiac application: the motion inside the myocardium being inhomogeneous, the computation of the histograms is done locally and the resulting term was added to the evolution equation proposed in [9].

Table 1 provides the mean and standard deviation of the error measures obtained for the complete dataset on the myocardial borders. We give for each measure (MAD , HD) the inter-observer distance IOD (first two columns), the error measures associated to the proposed method (columns 3 and 4), to Hamou's method (columns 5 and 6) and to Herbulot's

algorithm (two last columns). Fig. 1 shows tracking results for 2 sequences.

In parasternal short-axis view (first line of Table 1), it may be observed that our method provides small values for both criteria ($MAD = 1.31\text{mm}$ and $HD = 3.84\text{mm}$). Our segmentation results are thus close to the reference contours both on a global (MAD) and a local (HD) scale for this orientation. Moreover when compared to the corresponding IOD (two first columns in Table 1), it may be observed that the values obtained using the proposed method are very close to the inter-observer ones (1.09mm and 3.47mm for the MAD and HD respectively). This indicates that the segmentation provides consistent results in the sense that the difference with the experts reference is comparable to the distance between experts. This can also be seen from the standard deviation since the ones obtained with our algorithm are close to the ones from the experts. The same conclusions can be drawn for apical 4-chamber view, where the obtained errors (*e.g.* $MAD=1.39\text{mm}$) are even lower than the IOD ($MAD = 1.57\text{mm}$). We can also note that in MRI sequences, the computed errors are of the same order as for US data thus showing the ability of our algorithm to correctly track the myocardium in both US and MRI data.

The figures provided in the last four columns of Table 1 allow comparing the performance of the proposed method and two other motion based methods [1, 4]. The results show that our method yields better results (difference statistically significant at a level $p < 0.05$ for the Friedman rank test). Considering the error measures associated with our approach and Hamou's, they are both at least divided by 2 with our algorithm. This can be explained by the fact that our motion term is more accurate than the classical OF hypothesis applied to B-mode images. Similarly, comparing our results with Herbulot's, we can also note at least a 20% improvement of our results. This shows that our level conservation hypothesis is more efficient to drive the contour towards the true boundary than the separation of motion histogram proposed in [4].

6. CONCLUSION

In this article, we described a new motion prior energy that when minimized imposes a level consistency to the level-set function. This energy is then used for the segmentation and tracking of the myocardium. We evaluated the method on a dataset of 10 echocardiographic and 5 cine-MRI sequences (≈ 700 images) and showed a good agreement between our results and the experts references. It also compared favorably to two other recent methods [1, 4] with statistically significant differences.

7. ACKNOWLEDGMENT

We thank Dr. Ruta Jasaityte and Valérie Robesyn from Cardiovascular Imaging and Dynamics lab (Leuven) for providing the expert's contours. This work was partially supported by Agence Nationale de la Recherche US-Tagging

	IOD		Our method		Hamou [1]		Herbulot [4]	
	MAD	HD	MAD	HD	MAD	HD	MAD	HD
Parasternal short-axis	1.09 (0.38)	3.47 (1.16)	1.31 (0.91)	3.84 (2.07)	2.40† (0.97)	8.35† (3.21)	1.59† (0.56)	5.89† (2.40)
Apical 4-chamber	1.57 (0.52)	4.97 (1.44)	1.39 (0.48)	4.65 (1.44)	3.05† (1.26)	11.40† (5.06)	1.60† (0.50)	5.45† (1.62)
MRI	N/A	N/A	1.17 (0.31)	4.14 (1.15)	3.06† (1.30)	12.47† (4.80)	2.65† (0.90)	9.65† (3.70)

Table 1: Results of the segmentation of the whole myocardium. The results of our method and the one described in [1] and [4] as well as the Inter-Observers Distance (IOD) are shown in term of Mean Absolute Distance (MAD) and Hausdorff distance (HD). The given values are given in mm. † indicates that the difference was found significant ($p < 0.05$) when compared to our method.

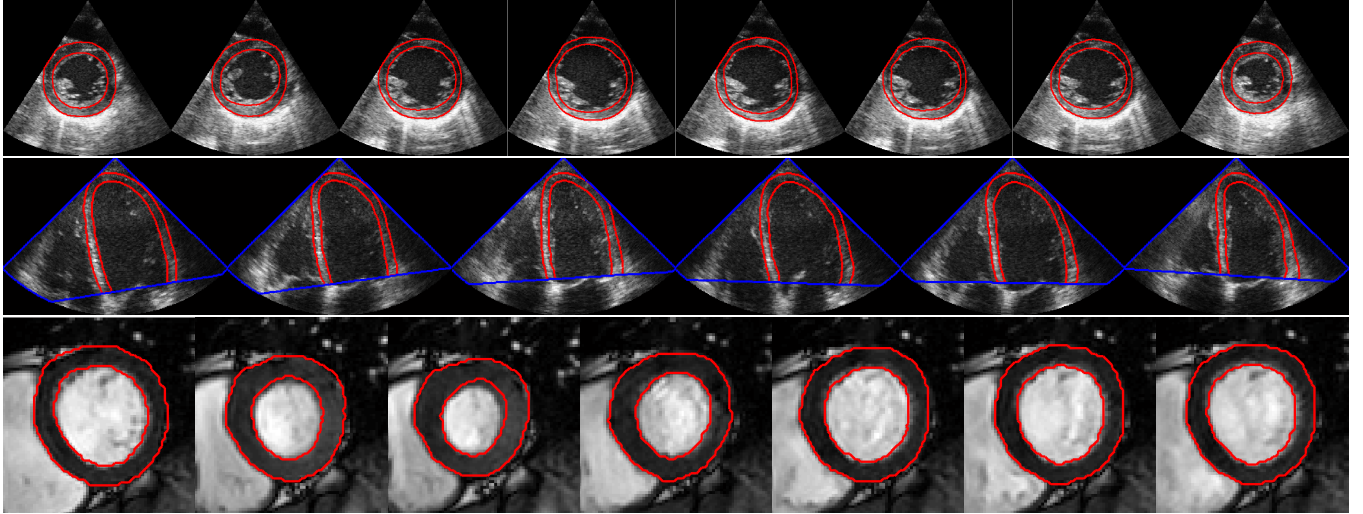


Fig. 1: Tracking results of our algorithm on a parasternal short-axis (top row), apical 4-chamber (middle row) and MRI (bottom row) sequence. Red: the tracking result; blue: the ROI.

grant), the region Rhône-Alpes (ExploraDoc and Accueil-Doc grants), Egide (PHC Tournesol grant) and the Fonds Wetenschappelijk Onderzoek - Vlaanderen (FWO) under the action Marie Curie Pegasus. This work was conducted in the framework of the LabEX PRIMES ("Physics Radiobiology Medical Imaging and Simulation").

8. REFERENCES

- [1] A. Hamou and M. El-Sakka, "Optical flow active contours with primitive shape priors for echocardiography," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, 2010.
- [2] N. Peterfreund, "Robust tracking of position and velocity with kalman snakes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 6, pp. 564–569, 1999.
- [3] T. Brox, A. Bruhn, and J. Weickert, "Variational motion segmentation with level sets," in *European Conference on Computer Vision (ECCV)*, Graz, Austria, Springer, LNCS, 2006, vol. 3951, pp. 471–483.
- [4] A. Herbulot, S. Jehan-Besson, S. Duffner, M. Barlaud, and G. Aubert, "Segmentation of vectorial image features using shape gradients and information measures," *J. of Math Imaging Vis.*, vol. 25, pp. 365–386, 2006.
- [5] Q. Zhang and R. Pless, "Segmenting cardiopulmonary images using manifold learning with level sets," in *Computer Vision for Biomedical Image Appl.*, 2005, vol. LNCS 3765, pp. 479–488.
- [6] D. Cremers, "Dynamical statistical shape priors for level set-based tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, pp. 1262–1273, 2006.
- [7] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer Verlag, 2002.
- [8] M. Alessandrini, A. Basarab, H. Liebgott, and O. Bernard, "Myocardial motion estimation from medical images using the monogenic signal," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1084–1095, 2013.
- [9] T. Dietenbeck, M. Alessandrini, D. Barbosa, J. D’hooge, D. Friboulet, and O. Bernard, "Detection of the whole myocardium in 2d-echocardiography for multiple orientations using a geometrically constrained level-set," *Medical Image Analysis*, vol. 16, pp. 386–401, 2012.
- [10] C. Fonseca, M. Backhaus, D. Bluemke, R. Britten, J. Chung, B. Cowan, I. Dinov, J. Finn, P. Hunter, A. Kadish, D. Lee, J. Lima, P. Medrano-Gracia, K. Shivkumar, A. Suinesiaputra, W. Tao, and A. Young., "The cardiac atlas project - an imaging database for computational modeling and statistical atlases of the heart," *Bioinformatics*, vol. 27, no. 16, pp. 2288–2295, 2011.